several low energy models (for example, 2-10 ), is(are) retained for a given target amino acid sequence. If desired, that model can then be used for various purposes, for example, to view the three-dimensional structure of the target amino acid sequence or by another computer program, *e.g.*, a program that can identify protein functional sites. A reduced model according to the invention can also be used to build more refined, or detailed, structural models, including heavy atom models and all-atom models.

Another aspect of the invention concerns computer programs that can convert an alignment of a target amino acid sequence with a template amino acid sequence into one or more three-dimensional reduced protein models comprising representations of side chains of amino acid residues comprising the probe amino acid sequence. In certain embodiments, such programs utilize at least one secondary constraint and one tertiary constraint for each side chain center of mass present in the probe amino acid sequence. In other embodiments, only some of the amino acid residues represented in the probe amino acid sequence have at least one tertiary and/or at least one secondary constraint that is acted on by the computer program. Embodiments of secondary constraints include those indicating the presence of a helix, and extended conformation, or anything else. Embodiments of tertiary constraints include positions in continuous three-dimensional space, positions lattice-based three-dimensional space, ranges of such positions, distances, ranges of distances, bond angles, ranges of bond angles, *etc.*

Embodiments of the invention that concern computer-assisted methods for determining a three-dimensional structure of a target amino acid sequence using a computer include those wherein the computer comprises a processor configured to receive and output data in accordance with executable code, *i.e.*, a program or computer control logic. Such methods include first inputting into the computer an alignment of a probe amino acid sequence with a template amino acid sequence. Then, by way of executable code, the processor is directed to produce from the alignment a three-dimensional reduced protein model comprised of representations

of side chains of amino acid residues comprising the target protein. This representation can then be output to an output device or to a storage device.

In preferred embodiments, the executable code comprises instructions for converting representations of the side chains of amino acid residues of the target protein to interaction centers (which can be represented as "beads" or pseudoatoms) connected by virtual covalent bonds. Each interaction center typically comprises a pseudoatom representing a center of mass of the side chain of the represented amino acid to which the interaction center corresponds, and each interaction center, except for the interaction centers representing the amino and carboxy terminal amino acid residues of the protein, is connected to an immediately proximal interaction center and an immediately distal interaction center via a virtual covalent bond to produce an interaction center chain. The program then projects the interaction center chain onto an underlying cubic lattice to produce a projected chain of interaction centers. In many embodiments, interaction centers have identity constraints associated therewith. Secondary constraints and/or tertiary constraints are then applied to a subset of, or all of, the interaction centers of the interaction center chain so as to produce a data set representing a three-dimensional model structure of the target protein. This method can further comprise iterating the foregoing steps. In each iteration, a different set of secondary and/or tertiary constraints can be applied to the interaction centers to produce a series of data sets representing three-dimensional model structures of the target protein. An energy computation can then be made for each member of the series of data sets. The data set(s) having the lowest computed energy(ies) are then preferably retained. Preferably, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 of the lowest energy data sets are retained or output to a data storage system to produce a stored data set. Alternatively, or in addition, one or more members of the data set can be output to an output device, such as a monitor on which the model can be visualized as a three-dimensional representation of the target protein. The member of the series of data sets having the lowest calculated energy can represent best, or highest quality, three-dimensional model structure of the target protein.

Definitions

The following terms have the following meanings when used herein and in the appended claims. Terms not specifically defined herein have their art recognized meaning.

As used herein, an "amino acid" is a molecule having the structure wherein a central carbon atom (the alpha (α)-carbon atom) is linked to a hydrogen atom, a carboxylic acid group (the carbon atom of which is referred to herein as a "carboxyl carbon atom"), an amino group (the nitrogen atom of which is referred to herein as an "amino nitrogen atom"), and a side chain group, R. When incorporated into a peptide, polypeptide, or protein, an amino acid loses one or more atoms of its amino and carboxylic groups in the dehydration reaction that links one amino acid to another. As a result, when incorporated into a protein, an amino acid is referred to as an "amino acid residue." In the case of naturally occurring proteins, an amino acid residue's R group differentiates the 20 amino acids from which proteins are synthesized, although one or more amino acid residues in a protein may be derivatized or modified following incorporation into protein in biological systems (*e.g.*, by glycosylation and/or by the formation of cystine through the oxidation of the thiol side chains of two non-adjacent cysteine amino acid residues, resulting in a disulfide covalent bond that frequently plays an important role in stabilizing the folded conformation of a protein, *etc.*). As those in the art will appreciate, non-naturally occurring amino acids can also be incorporated into proteins, particularly those produced by synthetic methods, including solid state and other automated synthesis methods. Examples of such amino acids include, without limitation, α-amino isobutyric acid, 4-amino butyric acid, L-amino butyric acid, 6-amino hexanoic acid, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norlensine, norvaline, hydroxproline, sarcosine, citralline, cysteic acid, t-butylglyine, t-butylalanine, phenylylycine, cyclohexylalanine, β-alanine, fluoro-amino acids,